CLAIMS:

1. A method of generating a text segmentation model for the segmentation of a text (100) into sections of text (102) on the basis of training data, wherein each section of text is assigned to a topic (108), the method of generating the text segmentation model comprising the steps of: ^

- generating a text emission model to provide a text emission probability being indicative of a section of text (102) being correlated to a topic (108),
- generating a topic sequence model to provide a topic sequence probability being indicative of a probability of a sequence of topics within the text,
- generating a topic position model to provide a topic position probability being indicative of a position of a topic (108) within the text (100),
- generating a section length model to provide a section length probability being indicative of the length of a section of text (102) that is assigned to a topic (108).

2. The method according to claim 1, wherein the training data comprises at least one text (100) segmented into sections of text (102), each section of text having assigned a topic (108).

3. The method according to claim 1 or 2, wherein the topic sequence model is adapted to account for a plurality of successive topic transitions by making use of a topic transition M-gram model.

4. The method according to any one of the claims 1 to 3, wherein the text emission probability is further determined with respect to the position of characteristic text portions within a section of text (102).

5. The method according to any one of the claims 1 to 4, wherein the text emission probability, the topic sequence probability, the topic position probability and the section length probability are determined with respect to a granularity parameter, influencing the number of sections (102) into which the text (100) is segmented.

6. A method of segmentation of a text (100) into sections of text (102) by making use of a text segmentation model being generated in accordance to any of the claims 1 to 5, the segmentation of the text being performed by selecting at least one probability of the group of probabilities consisting of: text emission probability, topic sequence probability, topic position probability and section length probability, and using the selected probabilities, the segmentation of the text further comprising assigning a topic (108) to each section of text (102).

7. The method according to claim 6, further comprising assigning a label (110, 112, 114) to each section of text, the label belonging to a set of labels (110, 112, 114) associated to the topic (108) being assigned to each section of text (102).

8. The method according to claim 6 or 7, wherein a granularity parameter influences the number of sections (102) into which the text (100) is segmented.

9. The method according to claim 7 or 8, further comprising:

- assigning a label (110, 112, 114) to a section (102) according to an ordered set of labels being associated to a topic (108)
- assigned to the section, assigning a label to a section (102) with respect to a text portion within the section, the text portion being characteristic for the label (110, 112, 114),
- assigning a label (110, 112, 114) to a section (102) with respect to a count statistics based on the training data, the count statistics being indicative

about a correlation probability between a topic (108) and the label (110, 112, 114).

10. The method according to any one of the claims 1 to 9, wherein modifications of the text emission probability, the topic sequence probability, the topic position probability and the section length probability are performed in response to a user's decision, the user having access to alter the text segmentation and assignment of topics (108) and labels (110, 112, 114) to sections of text (102).

11. A computer program product for the generation of a text segmentation model for the segmentation of a text (100) into sections of text (102) on the basis of annotated training data, wherein each section of text is assigned to a topic (108), the computer program product comprising program means for:

- generating a text emission model to provide a text emission probability being indicative of a section of text (102) being correlated to a topic (108),
- generating a topic sequence model to provide a topic sequence probability being indicative of a probability of a sequence of topics (108) within the text (100),
- generating a topic position model to provide a topic position probability being indicative of a position of a topic (108) within the text (100),
- generating a section length model to provide a section length probability being indicative of the length of a section of text (102) that is assigned to a topic.

12. The computer program product according to claim 11, wherein the topic sequence model is adapted to account for a plurality of successive topic transitions by making use of a topic transition M-gram model, and wherein the text emission probability is further determined with respect to the position of characteristic text portions within a section of text (102).

13. A computer program product for the segmentation of a text (100) into sections of text (102) by making use of a text segmentation model generated by a computer program product in accordance to claim 11 or 12, the computer program product for the segmentation of the text comprising program means for the segmentation of the text, the program means selecting at least one probability of the group of probabilities consisting of: text emission probability, topic sequence probability, topic position probability and section length probability, and using the selected probabilities, the program means being further adapted to assign a topic (108) to each section of text (102).

14. The computer program product according to claim 13, wherein a granularity parameter defines the number of sections (102) into which the text (100) is segmented.

15. The computer program product according to claim 13 or 14, further comprising program means being adapted to:

- assign a label (110, 112, 114) to a section (102) according to an ordered set of labels being associated to a topic (108) assigned to the section,
- assign a label (110, 112, 114) to a section (102) with respect to a text portion within the section, the text portion being characteristic for the label,
- assign a label (110, 112, 114) to a section (102) with respect to a count statistics based on the training data, the count statistics being indicative about a correlation probability between a topic and the label.

16. A computer program product according to any one of the claims 11 to 15, further comprising program means in order to perform modifications of the text emission probability, the topic sequence probability, the topic position probability and the section length probability in response to a user's decision, the user having access to alter the text segmentation and assignment of topics (108) and labels (110, 112, 114) to sections of text (102).

17. A computer system for the generation of a text segmentation model for the segmentation of a text (100) into sections of text (102) on the basis of annotated training data, wherein each section of text is assigned to a topic (108), the computer system comprising:

- means for generating a text emission model to provide a text emission probability being indicative of a section of text (102) being correlated to a topic (108),
- means for generating a topic sequence model to provide a topic sequence probability being indicative of a probability of a sequence of topics within the text,
- means for generating a topic position model to provide a topic position probability being indicative of a position of a topic within the text,
- means for generating a section length model to provide a section length probability being indicative of the length of a section of text (102) that is assigned to a topic.

18. The computer system according to claim 17, wherein the topic sequence model is adapted to account for a plurality of successive topic transitions by making use of a topic transition M-gram model, and wherein the text emission probability is further determined with respect to the position of characteristic text portions within a section of text (102).

19. A computer system for the segmentation of a text (100) into sections of text (102) by making use of a text segmentation model generated in accordance to claim 17 or 18 by a computer system, the computer system for the segmentation of the text comprising means being adapted to select at least one of the group of probabilities consisting of: text emission probability, topic sequence probability, topic position probability and section length probability, and using the selected probabilities, the

computer system means being further adapted to assign a topic (108) to each section of text (102).

20. The computer system according to claim 19, further comprising:

- means for assigning a label (110, 112, 114) to a section (102) according to an ordered set of labels being associated to a topic (108) assigned to the section,
- means for assigning a label (110, 112, 114) to a section (102) with respect to a text portion within the section, the text portion being characteristic for the label,
- means for assigning a label (110, 112, 114) to a section with respect to a count statistics based on the training data, the count statistics being indicative about a correlation probability between a text portion and the label.